



A data publication workflow tool

Deborah Smith

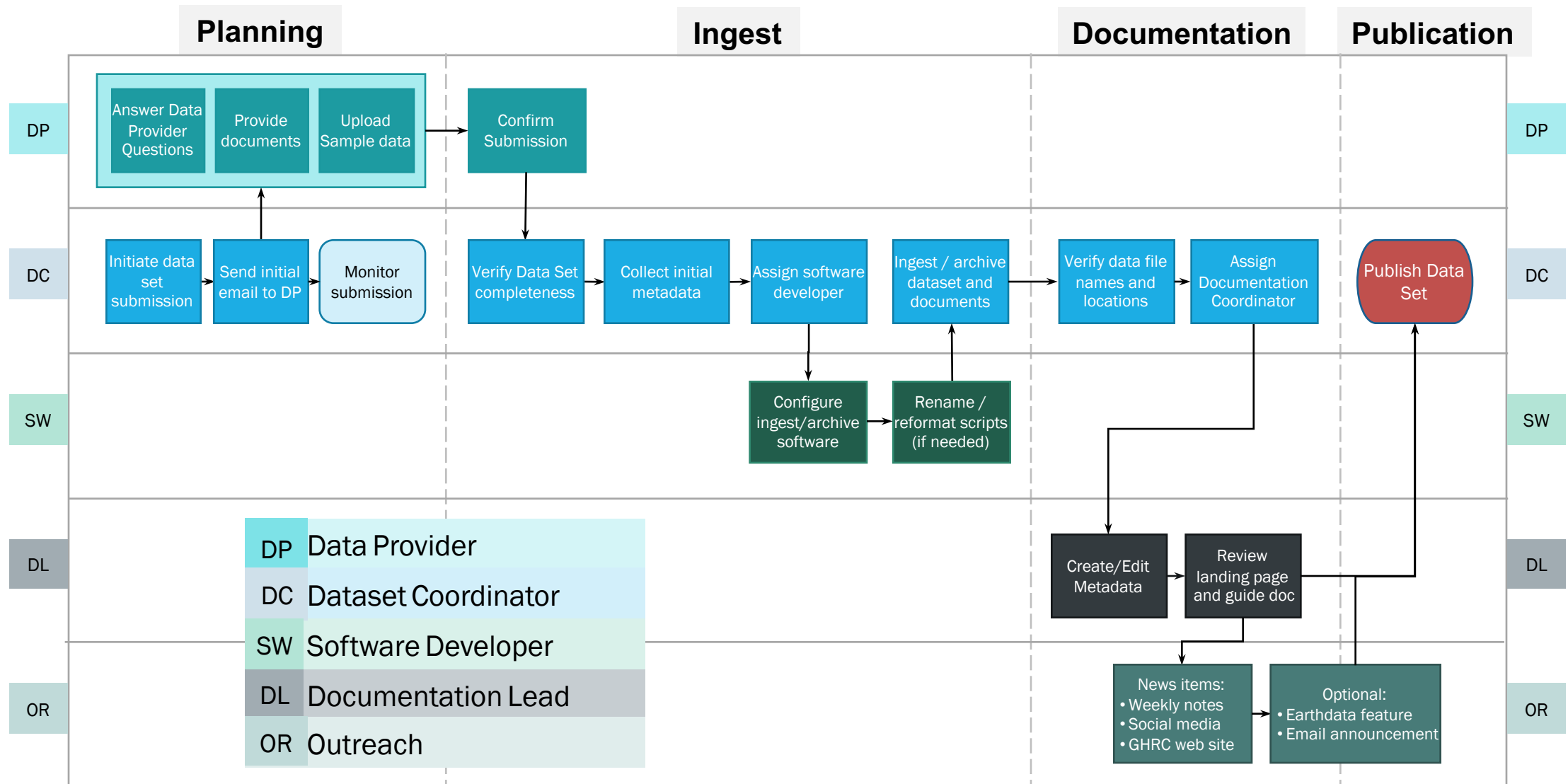
Ajinkya Kulkarni

2016 GHRC User Working Group Meeting
Sept 20-21, 2016



- This is a centralized data publication workflow tool for managing dataset publication at GHRC from acceptance of data from a provider through to publication to the end user.
- It is adapted from a tool (SAuS) originally developed at ORNL DAAC and shared with GHRC. We adapted the tool to meet our needs and upgraded it to work in a single coherent system based on Drupal 8.
- This tool provides a centralized system for Data Providers, the GHRC Data Management Group and the GHRC DAAC Managers. It automates many the steps required for dataset publication to reduce redundancy and repetition of effort.

GHRC Ingest Automation Swimlanes



* adapted from ORNL DAAC's Swimlanes

Key Components of the Data Workflow Tool

- Various data entry forms for data providers and GHRC data managers
- Views/dashboards for data managers and data providers
- Tracking and viewing of dataset workflow progress at each step
- Autofill information between forms
- Email notifications and email aggregation at one central place
- Integration with GCMD Keywords List (auto fetch/update keywords)
- REST API interface to GHRC catalog database
- Management of users and user permissions
- Management of DOI and link management (edit, validate, submit, etc.)

For the Data Provider

Forms

- Archival Interest Form
- Questionnaire Form

Views

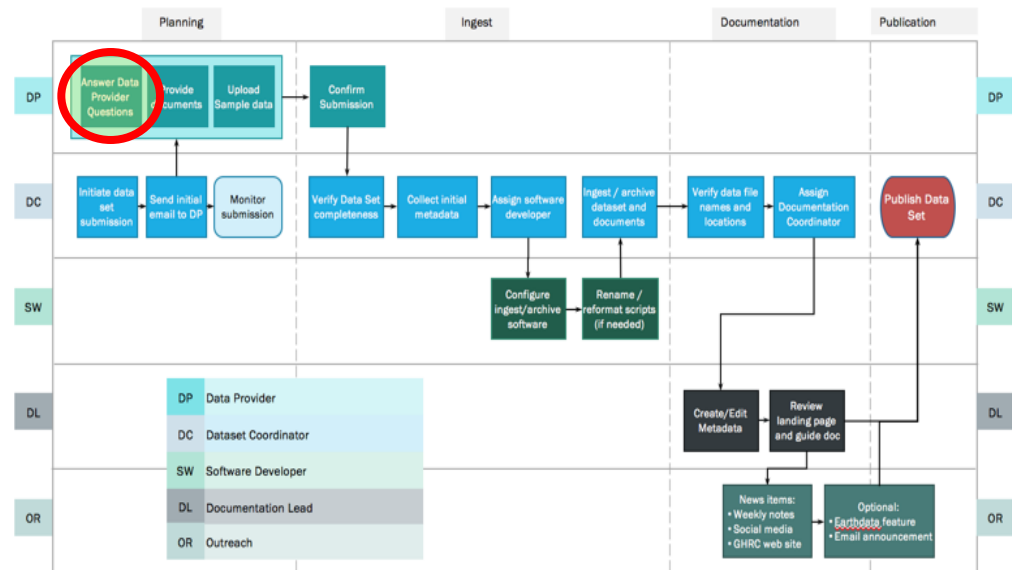
- Check publication status of own data
- Edit/clone own records

For the GHRC Data Management Group

- Metadata Records
- User Guides
- Instruments
- Platforms
- Organizations
- Projects
- Publications Tracking
- Track publication status of all datasets
- View email communications
- Quickly link to various entries and submissions

Archival Interest Form

Step 1: A data provider submits a request for submitting data to GHRC for archival



Title *

PI OF DATASET

First Name *

Middle Name

Last Name *

Email *

Organization *

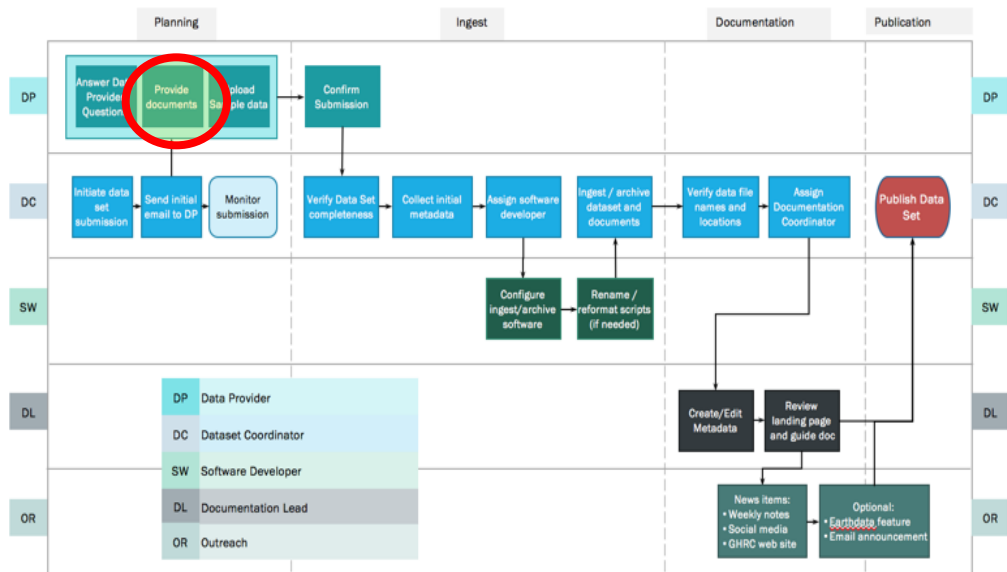
Description of your data *

Estimated Data Volume(GB) *

Science value of the dataset *

Data Provider Questionnaire Form

Step 2: If accepted for the archival at GHRC, the data provider fills out questionnaire form



Title *

Information about your dataset

Dataset description

Temporal and spatial characteristics

Data preparation and delivery

Have you looked at our recommendations for the preparation of data files and documentation? *

☐ No

☐ Yes

<https://ghrc.nsstc.nasa.gov/home/ghrc-docs/data-management>

[Show row weights](#)

WHO PRODUCED THIS DATASET?

First Name *

Middle Name

Last Name *

Email *

Organization *

Department *

Role *

- Select a value -

Order of name in citation

#edit-group-dataset-information on this page

Step 3: GHRC personnel are then automatically assigned to publication workflow tasks and email notifications are sent out

Ingest Coordinator

Marilyn (11)

Science Coordinator

Deborah (14)

Software Coordinator

Matt (12)

Database Coordinator

Mary (13)

OPS Coordinator

Lamar (9)

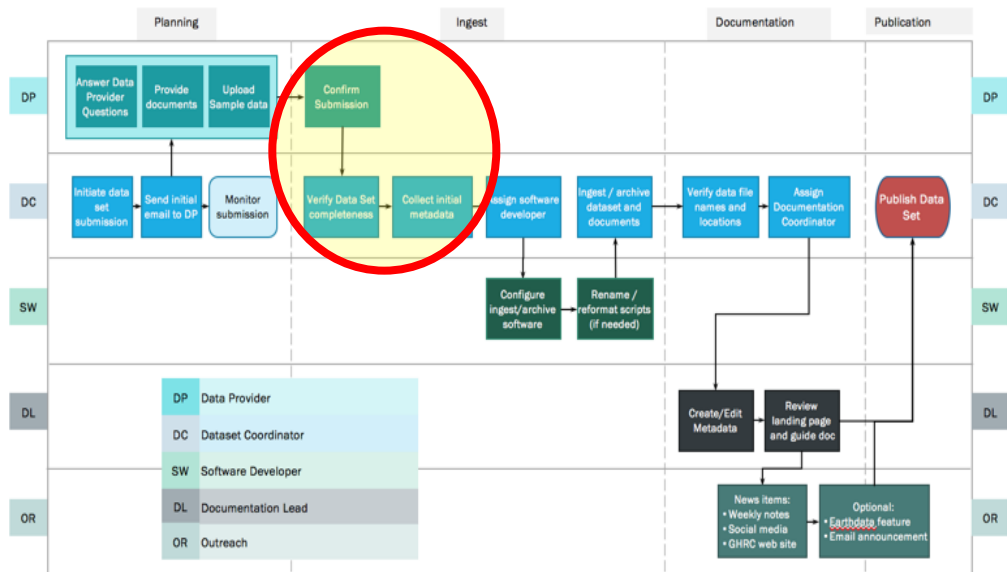
GIBS Coordinator

Shannon (10)

Tools Coordinator

ajinkya (1)

Step 4: GHRC Data Management Group lead creates an initial metadata record



Home » Node » Add content

Data Identification

Acquisition and Science Keyword Information

Spatial/Temporal Information

Distribution Information

DOI / Citation

Housekeeping

Dataset Longname *

Dataset Shortname

Dataset Description

Dataset Version

Dataset Processing Level

- None - ▼

[Show row weights](#)

DATA PRODUCER(PI)/TECHNICAL CONTACT(CO-PIS)

First Name *

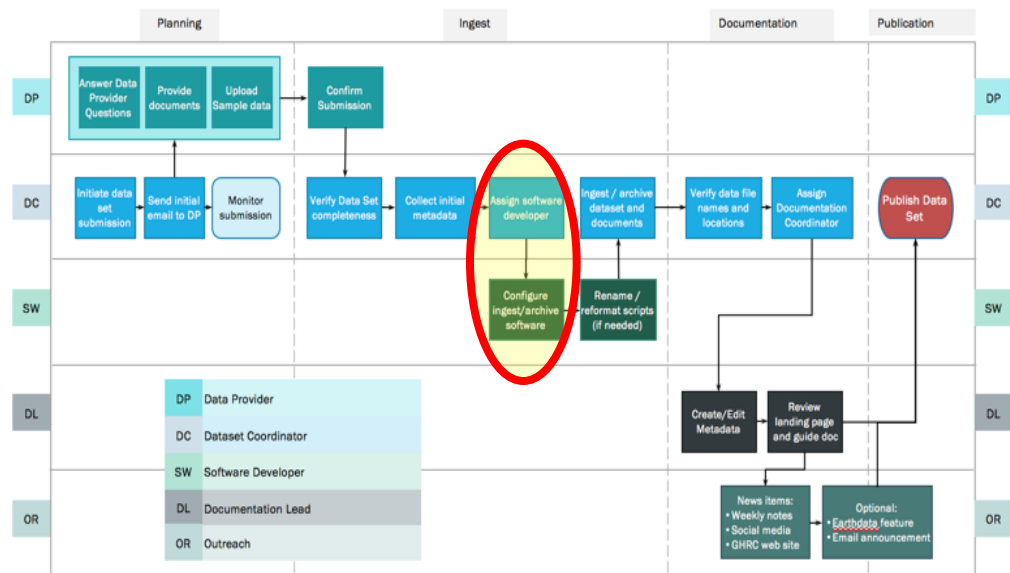
Middle Name

Last Name *

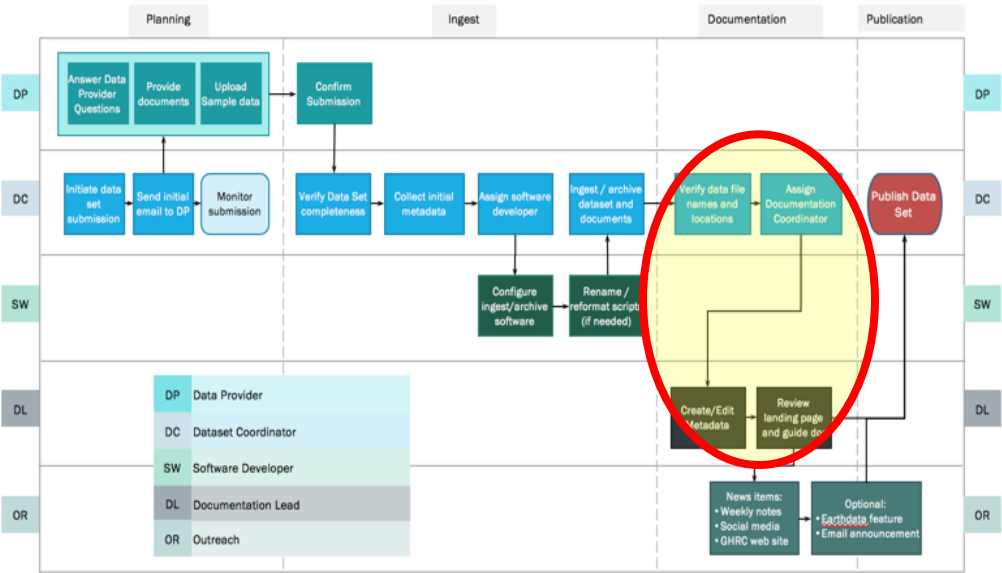
Email *

Organization *

Step 5: A software developer works on configuring ingest/archival software



Step 6: A science team works on the documentation and completes the metadata record



Title *

Introduction

Citation

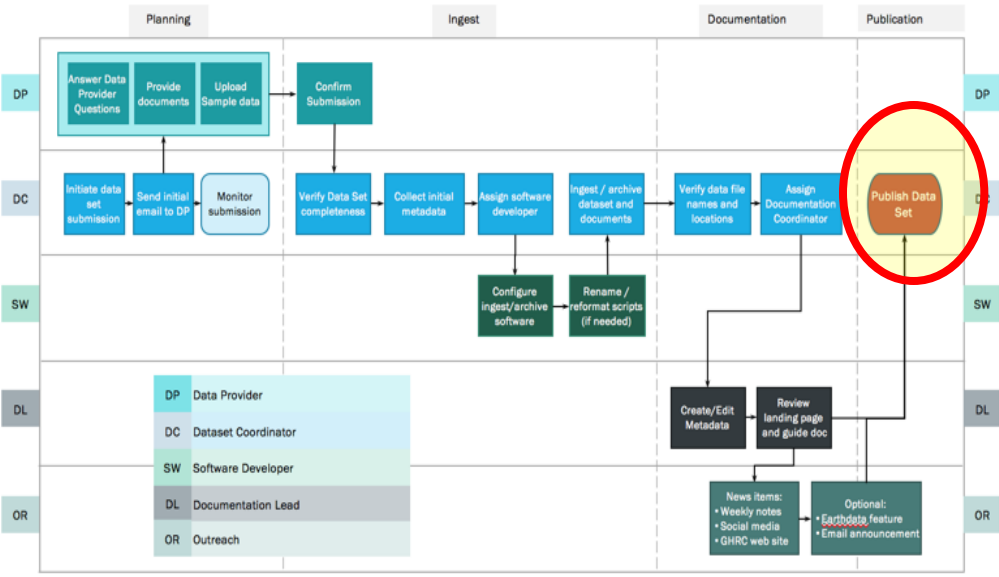
Keywords

Doppler Velocity, Radar Reflectivity

Campaign

Instrument Description

Step 7: An ingest coordinator uses the tracker to check the progress and decides when to publish the dataset



Title *

Start Date

End Date

PI Name

Status

Assignments

Initiated

Metadata Record

Processing

Catalog

Documentation

Publish

Accepted for Archival?

☐ N/A

☒ No

☐ Yes

Data Provider Questions

☐ N/A

☒ No

☐ Yes

Initiated

☐ N/A

☒ No

☐ Yes

Metadata Record

☐ N/A

☒ No

☐ Yes

Processing

☐ N/A

Dataset Publication Statuses

GHRC DAAC Managers can use the central view to track the progress of all datasets currently in the workflow process

Filters:

Published status Initiated Cataloged Processing Documentation Metadata Record

Title	Data Producer (PI)	Authored on	Initiated	Metadata	Processing	Catalog	Documentation	Published
Hurricane and Severe Storm Sentinel (HS3) Hurricane Imaging Radiometer (HIRAD)	Cecil Daniel	2016-09-13	Yes	Yes	Yes	Yes	Yes	Yes
Hurricane and Severe Storm Sentinel (HS3) High-altitude Imaging Wind and Rain Airborne Profiler (HIWRAP)	Gerald Heymsfield	2016-09-13	Yes	No	No	No	No	No
Hurricane and Severe Storm Sentinel (HS3) Global Hawk High Altitude MMIC Sounding Radiometer (HAMSR)	Bjorn Lambrigtsen	2016-09-13	Yes	Yes	No	No	No	No

























A centralized email logging system allows for Data Management Group and DAAC Managers to view the external as well as internal communications for any particular dataset

Email Communications

Subject	Date	From	To	Attachments
Regarding the SAuS Activity	2016-09-13 20:47:21	GHRC SAuS System	ghrc@ghrc.nasa.gov	
Regarding the SAuS Activity	2016-09-13 20:47:21	GHRC SAuS System	ghrc@ghrc.nasa.gov	
Regarding the SAuS Activity	2016-09-13 20:45:56	GHRC SAuS System	ghrc@ghrc.nasa.gov	
Regarding the SAuS Activity	2016-09-13 20:47:21	GHRC SAuS System	ghrc@ghrc.nasa.gov, ghrc@ghrc.nasa.gov	
Regarding the SAuS Activity	2016-09-13 23:25:17	GHRC SAuS System	ghrc@ghrc.nasa.gov, ghrc@ghrc.nasa.gov	
Regarding the SAuS Activity	2016-09-13 23:25:17	GHRC SAuS System	ghrc@ghrc.nasa.gov	
Regarding the SAuS Activity	2016-09-13 23:25:17	GHRC SAuS System	ghrc@ghrc.nasa.gov	

We make use of Drupal taxonomy to manage various GCMD keywords, tags, and to autofill values at a centralized place

[Show row weights](#)

NAME	OPERATIONS
 Earth Remote Sensing Instruments	Edit 
 Active Remote Sensing	Edit 
 Altimeters	Edit 
 Lidar/Laser Altimeters	Edit 
 AIRBORNE LASER SCANNER	Edit 
 ATM..Airborne Topographic Mapper	Edit 
 CALIOP..Cloud-Aerosol Lidar with Orthogonal Polarization	Edit 
 GLAS..Geoscience Laser Altimeter System	Edit 
 LIDAR ALTIMETERS..Light Detection and Ranging Altimeters	Edit 
 LVIS-Camera..Digital Mapping Camera paired with LVIS	Edit 
 LVIS-GH..Land, Vegetation, and Ice Sensor - Global Hawk	Edit 
 LVIS..Land, Vegetation, and Ice Sensor	Edit 


 [Link Types](#)


 [Locations](#)


 [Platforms](#)

 [Relationship to GHRC](#)

 [Science and Services Keywords](#)

 [Short Names](#)

 [Spatial Coverage](#)

 [Spatial Resolution](#)

 [Sponsor/Project](#)

* retrieved from http://gcmd.nasa.gov/learn/keyword_list.html

- Development version available - <https://ghrcdev.nsstc.nasa.gov/data-publication/>
- Demo User Login Information
 - User : ghrc-demo-user
 - Password: ghrc-demo-user
- Once logged in, check out the Archival Interest Form and the PI Questionnaire Form.
- Both links to these forms located on <https://ghrcdev.nsstc.nasa.gov/data-publication/>

Anticipated Release Schedule

- Test version – 10/30/2016
- Operational version – 3/31/2017



Discussion

1. Has any terminology been used in this form that you do not understand? If so, please explain.
2. Do you feel any pertinent dataset information is missing from this form? If so, please explain.

2016 GHRC User Working Group Meeting
Sept 20-21, 2016

